Project Proposal

Christophe

October 17, 2015

```
# load data
library(RCurl)
url <- paste(url,"STD_Rate_Data_File.csv", sep = '')
std.data <- getURL(url)
std.data <- read.csv(textConnection(std.data))
df <- data.frame(std.data)
df <- subset(df, df$Race.Ethnicity != 'Unknown' & df$Gender.Code != 'U')
df$STD.Cases <- as.numeric(df$STD.Cases)
df$Rate <- as.numeric(df$Rate)
display_variables <- c("Disease", "State", "Gender",
                       "Race.Ethnicity", "STD.Cases","Population","Rate")

display_data <- df[display_variables]
kable(head(display_data, n = 15))
```

	Disease	State	Gender	Race.Ethnicity	STD.Cases	Population	Rate
1	Chlamydia	Alabama	Female	American Indian or Alaska Native	242	14724	183
2	Chlamydia	Alabama	Female	Asian or Pacific Islander	388	34099	116
3	Chlamydia	Alabama	Female	Black or African American	485	687266	159
4	Chlamydia	Alabama	Female	Hispanic	86	88045	104
5	Chlamydia	Alabama	Female	White	271	1659332	189
7	Chlamydia	Alabama	Male	American Indian or Alaska Native	411	14169	20
8	Chlamydia	Alabama	Male	Asian or Pacific Islander	251	30262	59
9	Chlamydia	Alabama	Male	Black or African American	337	599224	358
10	Chlamydia	Alabama	Male	Hispanic	319	107987	13
11	Chlamydia	Alabama	Male	White	55	1586915	39
19	Chlamydia	Alaska	Female	American Indian or Alaska Native	160	58271	279
20	Chlamydia	Alaska	Female	Asian or Pacific Islander	181	27682	363
21	Chlamydia	Alaska	Female	Black or African American	186	14180	167
22	Chlamydia	Alaska	Female	Hispanic	38	21358	308
23	Chlamydia	Alaska	Female	White	62	228720	311

Research question

Are there disparities in Chlamydia incidences when accounting for ethnicity and state/region in the US in 2013?

Cases

The number of cases and disease incidence rates are reported by year, gender of patient, race/ethnicity, type of STD, and state.

To investigate this research question we are only taking cases reported for the year 2013.

```
cases_table <- aggregate(list('Cases in 2013' = df$STD.Cases),
                by = list(Ethnicity = df$Race.Ethnicity, Gender = df$Gender), FUN=sum)
cases_table <- cases_table[order(cases_table$Ethnicity),]
kable(cases_table)
```

	Ethnicity	Gender	Cases.in.2013
1	American Indian or Alaska Native	Female	13043
6	American Indian or Alaska Native	Male	12887
2	Asian or Pacific Islander	Female	11048
7	Asian or Pacific Islander	Male	14660
3	Black or African American	Female	10714
8	Black or African American	Male	15617
4	Hispanic	Female	14371
9	Hispanic	Male	11796
5	White	Female	11831
10	White	Male	13710

```
kable(aggregate(list('Cases in 2013' = df$STD.Cases),
                by = list(State = df$State), FUN=sum))
```

State	Cases.in.2013
Alabama	2845
Alaska	2582
Arizona	2652
Arkansas	2173
California	2270
Colorado	1923
Connecticut	3403
Delaware	2564
District of Columbia	2085
Florida	2605
Georgia	2315
Hawaii	2050
Idaho	2724
Illinois	2053
Indiana	2613
Iowa	2195
Kansas	3148
Kentucky	2318
Louisiana	2212
Maine	2901
Maryland	2495
Massachusetts	1699
Michigan	2674

State	Cases.in.2013
Minnesota	2497
Mississippi	3254
Missouri	2918
Montana	3081
Nebraska	2817
Nevada	2138
New Hampshire	2735
New Jersey	2859
New Mexico	1664
New York	2371
North Carolina	2910
North Dakota	3097
Ohio	2040
Oklahoma	1960
Oregon	2610
Pennsylvania	1914
Rhode Island	2466
South Carolina	2949
South Dakota	2672
Tennessee	3152
Texas	3023
Utah	2047
Vermont	2100
Virginia	3102
Washington	2517
West Virginia	2620
Wisconsin	2790
Wyoming	2875

Data collection

Beginning in 2003, all 50 states and the District of Columbia had converted from summary hardcopy reporting to electronic case-specific reporting via NETSS (with the exception of congenital syphilis, which is still reported on hardcopy forms by several areas). Also, before 1996, Chlamydia reporting was voluntary, and thus sporadic. From 1995 - 2000, upstate New York did not report Chlamydia, and thus the total US denominator population for Chlamydia excludes the New York state population for the years 1995 - 2000. However, New York City did report Chlamydia before the year 2000, and the New York City population and Chlamydia cases are included for 1984 - 2000. (Sexually Transmitted Disease Morbidity by Race and Age. (n.d.). Retrieved October 17, 2015.)

It is unclear how many layers of abstraction exists, it appears that the information is compiled from provider level reports of each patient, then aggregated by the various health agencies and reported to CDC. However, it could be the case that each patient or incidence is reported to the CDC and the CDC does the aggregation.

Type of study

This study is an observational study on Sexually Transmitted Disease (STD) morbidity case records reported to the National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention (NCHHSTP), Centers for Disease Control and Prevention (CDC).

Data Source

US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for HIV, STD and TB Prevention (NCHSTP), Division of STD/HIV Prevention, Sexually Transmitted Disease Morbidity 1996 - 2013, by gender, age group and race/ethnicity, CDC WONDER Online Database.

CDC WONDER Online Database Query Date: Oct 17, 2015 8:30:27 AM

Response

The response variable or our variable of interest is the rate of STD (specifically Chlamydia). The rate of STD cases is measured as a numerical variable and at incidence per 1,000 people.

Explanatory

There are three explanatory variables in the data set, they are the ethnicity, gender, and state. Ethnicity, gender, and state are considered to be categorical variables.

Relevant summary statistics

Note that the sample size is 51 for all ethnicities because the observations are for all 50 states and the District of Columbia for 2013.

```
sum_stat_m <- aggregate(list(Rate = df$Rate), by = list(Ethnicity = df$Race.Ethnicity,
    State = df$State), FUN = sum)

sum_stat_m <- describeBy(sum_stat_m$Rate, sum_stat_m$Ethnicity, mat = TRUE)
rownames(sum_stat_m) <- sum_stat_m$group1
sum_stat_m <- sum_stat_m[-(5), -(1:3)]
sum_stat_m <- round(sum_stat_m, 2)
kable(sum_stat_m, align = "c")
```

	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
American Indian or Alaska Native	51	376.82	150.15	381	375.61	185.32	121	760	639	0.12	-0.69	21.02
Asian or Pacific Islander	51	194.73	126.45	152	184.15	123.06	13	605	592	0.83	0.43	17.71
Black or African American	51	508.49	122.85	501	509.34	78.58	289	737	448	-0.01	-0.83	17.20
Hispanic	51	421.08	132.93	432	433.32	163.09	82	592	510	-0.70	-0.32	18.61
White	51	258.43	95.99	276	262.61	81.54	30	544	514	-0.18	0.53	13.44